

Group Members: Alperen Duyan

Zeynep Menteshoglu

Project Background and Motivation

In the field of education, it is necessary to know what students learned after teaching each different concept. Most common way of measuring what they learned is to have an exam. As technology and relevant tools are becoming more common in each field, it has started to be used in education, too. One of the ways in which technology is commonly used to assess students is “Computerized Adaptive Testing (CAT),” which is based on a mathematical modeling that calculates the ability of a learner after answering each question in the background. This calculation is used to make a choice about the next question learner will face. When a student answers an item coming from item bank correctly, his/her ability increases, and s/he faces harder questions compared to the one s/he answered previously. In this way, learners can be measured individually based on their personal learning goals rather than making comparison with other learners taking the same exam. Therefore, CAT is commonly used to provide personal learning pathways. In CAT, item bank includes many different questions specific to different ability levels. These questions are prepared by experts in the field, and it is extremely time consuming. In addition to preparing questions, pilot studies are completed to decide the ability level of each item. However, artificial intelligence (AI) started to take different roles in each field and AI related tools can also be used to enhance the quality of education in different ways. Therefore, we decided to create an AI tool which can be used by students’ themselves or teachers to create an exam that is prepared by taking care of CAT principles such as enhancing the difficulty of an item with each correct answer. This project will also enable students to assess themselves without going to the test center when they are at home, for instance, and it was one of other motivations that led us to develop this tool.

Overview of System

In this project, we have three main blocks with the following functions;

1. **Speech to text Block:** The aim of this module is to take a voice input from the student about the topic s/he wants to take a quiz on and transform it to text to be able to send this as an input to LLM to generate questions and images.
2. **Question and Image Prompt Generation:** This module takes the input from Speech to text and creates a question for us. Our system prompt is as follows,

"role": "system",

"content": "I am preparing an educational quiz. I will provide you with a topic and you will give me questions. At first message, just wait for me to give you the topic. "

"I want you to ask me one question at a time. I want a difficulty scale from 1 to 7 where 1 is the easiest and 7 is the hardest. "

"I want you to start with a question with difficulty level 3. Each time I give a correct answer, increase the difficulty by 1, and if difficulty is 7, do not change it. "

"Each time I give a wrong answer, decrease the difficulty by 1. If difficulty is 1, do not change it. "

"I will use these questions in an app, so I want you to give me the question as a Python dictionary. "

"The format is like this: {'Question': '', 'Options': [], 'Answer': '', 'ImagePrompt': ''}. "

"Do not include any comments, just give me the formatted question."

This prompt has the instructions for LLM about the formatting of the question. We implemented our code by considering the question data is a python dictionary which has Question, Answer, Options, Image prompt, difficulty level. In our prompt we stated that we have 7 difficulty levels, and we started to have questions with difficulty level 3 to start with an average difficulty question. This block also creates an image prompt for the image generation block.

3) Image Generation: This block takes the image prompt from the image prompt generation and creates an image about the question. With this way, we wanted to make students more engaged to the learning material and make our quiz more inclusive by having visual aspects as well.

Overall, when the program is run, it starts with saying the topic to the tool blocks 2 and 3 continuously run in a loop.

Technical Aspects

In this project, we used “*Whisper-1*, *ChatGPT gpt-4o-mini*, and *DALL-E-2*” respectively. “Whisper” was the tool we integrated into our project as we were using the library Openai API to transform speech into the text. Next, “*ChatGPT*” was used as again we were using OpenAI API. The most time-consuming process in our game is the image generation process. Time required for this model is not that important when we compare it to time consuming of image generation. Therefore, to get a more accurate output, we used this model instead of the lightweight LLM model. For the image generation we used *DALL-E-2* as that is one of the lightest image generation models in OpenAI API since image generation is the most time-consuming event in our program.

During the development process, we benefited from the class notes especially for “openai api” functions and methods. For some syntax errors we asked help from ChatGPT. We used python pygame library to develop our quiz program and integrate our tools.

Discussion

We were a group of just two people so we could have the opportunity to work on all the modules together. We brainstormed together and came together regularly to make our thoughts come true.

Some ethical considerations related to this project could be that: data privacy could be an issue while using the tool as the system needs to anonymize the data personal data, language can create some barriers as the proficiency of language can affect student performance, this tool can create questions that reflects some biases in different issues like gender, and it can be a disadvantage for visually impaired learners to have images for questions.

The CAT model supports learners to have their own individualized pathways and causing personalized interactions between students and educational tools. Moreover, speech-to-text component improves the tool’s accessibility for diverse groups of learners. In addition, automating quiz generation using AI tools will cause educators to save time and effort, enabling them to focus on different tasks. In the future, we expect humans to be at the center and

technology is used to enhance human potential rather than replace it. People will also need to navigate AI-integrated environments, and this project can be helpful to prepare students for the future by showing how people can collaborate with AI tools.

In the future, the AI models used in this project are expected to be enhanced. For instance, *Whisper* can start offering better support by handling diverse accents and noisy environments. Moreover, these tools will be faster gradually and response time to get response will decrease. If this tool is recreated in 1-3 years, our tool can start creating animations instead of images for quiz questions, it can support multiple languages effectively, it can also provide feedback about the progression of each student by including some analytics. Furthermore, it could be harder to protect data misuse as AI progresses very quickly and may need continuous updates because of rapid advancements for maintaining necessities.